

HOMOLOGOUS PROTEIN:

Top 10 BLAST Hits:

	Score	E
	(bits)	Value
Sequences producing significant alignments:		
CRA 18000005098398 /altid=gi 4885535 /def=ref NP_005437.1 puri...	857	0.0
CRA 335001098681202 /altid=gi 11417813 /def=ref XP_009854.1 pu...	857	0.0
CRA 1000682348238 /altid=gi 6469324 /def=gb AAF13303.1 AF065385...	855	0.0
CRA 18000005129684 /altid=gi 6754966 /def=ref NP_035158.1 puri...	621	e-177
CRA 18000005027891 /altid=gi 6981322 /def=ref NP_036853.1 p2X6...	604	e-172
CRA 148000001425983 /altid=gi 7920253 /def=gb AAF70599.1 AF2050...	360	2e-98
CRA 18000005038217 /altid=gi 7447773 /def=pir S71344 purinergi...	348	8e-95
CRA 18000005027890 /altid=gi 1709522 /def=sp P51578 P2X5_RAT P2...	345	7e-94
CRA 18000005064403 /altid=gi 4505549 /def=ref NP_002551.1 puri...	318	9e-86
CRA 18000005196095 /altid=gi 4099121 /def=gb AAD00553.1 {U8399...	318	9e-86

EST:

	Score	E
	(bits)	Value
Sequences producing significant alignments:		
gi 11617343 /dataset=dbest /taxon=96...	1164	0.0
gi 6992441 /dataset=dbest /taxon=960...	648	0.0
gi 4990980 /dataset=dbest /taxon=9606 ...	579	e-163
gi 10325489 /dataset=dbest /taxon=96...	464	e-128
gi 2195075 /dataset=dbest /taxon=9606 ...	287	4e-75

EXPRESSION INFORMATION FOR MODULATORY USE:

gi|11617343 Brain- anaplastic oligodendroglioma
gi|6992441 Chronic lymphocytic leukemia
gi|4990980 Lung- carcinoid
gi|10325489 lung - large cell carcinoma
gi|2195075 Colon

Tissue expression:

Whole brain

TOP10 = 5600000

```

1  MGSPGATTGW  GLLDYKTEKW  ALLAKKGYYE  RDLEPQFSII  TKLKGVSVTQ
51 IKELGNRLWD  VADFKVPPQG  ENVFFLVTFN  LVTPAQVQGR  CPEHPSVPLA
101 NCWVDEDCPE  GEGGTHSHGV  KTGQCQVFNG  THRTCEIWSW  CPVESGVVPS
151 RPLLAQAQNF  TLFIKNTVTF  SKFNFSKSNA  LETWDPTYFK  HCRYEPQFSP
201 YCPVFRIGDL  VAKAGTFED  LALLGGSVGI  RVHWDCDLDT  GDSGCWPHYS
251 FQLQEKSYNF  RTATHWWEQP  GVEARTLLKL  YGIRFDILVT  GQAGKFGLIP
301 TAVTLGTGAA  WLGVTFFCD  LLLLYVDREA  HFYWRKYEE  AKAPKATANS
351 VWRELALASQ  ARLAECLRRS  SAPAPTATAA  GSQTQTPGWP  CPSSDTHLPT
401 HSGSL (SEQ ID NO: 2)

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 3

```

1      129-132 NGTH
2      159-162 NFTL
3      174-177 NFSK

```

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

368-371 RRSS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 2

```

1      17-19 TEK
2     131-133 THR

```

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 2

```

1     217-220 TFED
2     336-339 TKYE

```

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 10

```

1      2-7 GSPGAT
2      5-10 GATTGW
3     45-50 GVSVTQ
4     113-118 GGTHSH
5     119-124 GVKTGQ
6     130-135 GTHRTC
7     146-151 GVVPSR
8     225-230 GGSVGI
9     297-302 GLIPTA
10    306-311 GTGAAW

```

[6] PDOC00932 PS01212 P2X_RECEPTOR
ATP P2X receptors signature

225-251 GGSVGIRVHWDCDLDTGDSGCWPHYSF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	69	89	0.782	Putative
2	299	319	1.835	Certain

BLAST Alignment to Top Hit:

```

>CRA|18000005098398 /altid=gi|4885535 /def=ref|NP_005437.1|
purinergic receptor P2X-like 1, orphan receptor; P2X
specifically expressed in skeletal muscle; purinoceptor
P2X6 [Homo sapiens] /org=Homo sapiens /taxon=9606
/dataset=nraa /length=431

```

Length = 431
Score = 857 bits (2189), Expect = 0.0
Identities = 405/431 (93%), Positives = 405/431 (93%), Gaps = 26/431 (6%)

Query: 1 MGSPGATTGWGLLDYKTEK-----WALLAKKGQERDLE 34
MGSPGATTGWGLLDYKTEK WALLAKKGQERDLE
Sbjct: 1 MGSPGATTGWGLLDYKTEKYVMTRNWRVGAQRLQLQFGIVVYVVGWALLAKKGQERDLE 60

Query: 35 PQFSIITKLKGVSVTQIKELGNRLWDVADFVKPPQGENVFFLVNINFLVTPAQVQGRCPHEH 94
PQFSIITKLKGVSVTQIKELGNRLWDVADFVKPPQGENVFFLVNINFLVTPAQVQGRCPHEH
Sbjct: 61 PQFSIITKLKGVSVTQIKELGNRLWDVADFVKPPQGENVFFLVNINFLVTPAQVQGRCPHEH 120

Query: 95 PSVPLANCWVDEDCPEGEGGTHSHGVKTGQCVVFNGTHRTCEIWSWCPVESGVVPSRPLL 154
PSVPLANCWVDEDCPEGEGGTHSHGVKTGQCVVFNGTHRTCEIWSWCPVESGVVPSRPLL
Sbjct: 121 PSVPLANCWVDEDCPEGEGGTHSHGVKTGQCVVFNGTHRTCEIWSWCPVESGVVPSRPLL 180

Query: 155 AQAQNFTLFIKNTVTFSKFNFSKSNALETWDPTYFKHCRYEPQFSPYCPVFRIGDLVAKA 214
AQAQNFTLFIKNTVTFSKFNFSKSNALETWDPTYFKHCRYEPQFSPYCPVFRIGDLVAKA
Sbjct: 181 AQAQNFTLFIKNTVTFSKFNFSKSNALETWDPTYFKHCRYEPQFSPYCPVFRIGDLVAKA 240

Query: 215 GGTFFEDLALLGGSVGIRVHWDCLDLDGSGCWPHYSFQLQEKSYNFRATATHWWEQPGVEA 274
GGTFFEDLALLGGSVGIRVHWDCLDLDGSGCWPHYSFQLQEKSYNFRATATHWWEQPGVEA
Sbjct: 241 GGTFFEDLALLGGSVGIRVHWDCLDLDGSGCWPHYSFQLQEKSYNFRATATHWWEQPGVEA 300

Query: 275 RTLLKLYGIRFDILVTGQAGKFGLIPTAVTLGTGAAWLGVVTFCDLLLLLYVDREAHFYW 334
RTLLKLYGIRFDILVTGQAGKFGLIPTAVTLGTGAAWLGVVTFCDLLLLLYVDREAHFYW
Sbjct: 301 RTLLKLYGIRFDILVTGQAGKFGLIPTAVTLGTGAAWLGVVTFCDLLLLLYVDREAHFYW 360

Query: 335 RTKYEEAKAPKATANSVWRELALASQARLAECRLRSSAPAPTATAAGSQQTPTGWPCPSS 394
RTKYEEAKAPKATANSVWRELALASQARLAECRLRSSAPAPTATAAGSQQTPTGWPCPSS
Sbjct: 361 RTKYEEAKAPKATANSVWRELALASQARLAECRLRSSAPAPTATAAGSQQTPTGWPCPSS 420

Query: 395 DTHLPTHSGSL 405
DTHLPTHSGSL
Sbjct: 421 DTHLPTHSGSL 431 (SEQ ID NO: 4)

Hammer search results (Pfam):
Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
CE00369	E00369 P2X6_receptor	1180.5	0	2
PF00864	ATP P2X receptor	870.0	7.4e-258	1
CE00207	CE00207 PURINERGIC	366.8	5.9e-111	1
CE00370	E00370 P2X4_receptor	336.8	1.9e-109	1
CE00368	E00368 P2X7_receptor	124.1	6.5e-36	1
PF00095	WAP-type (Whey Acidic Protein) 'four-disulfi	8.7	1.1	1
PF01841	Transglutaminase-like superfamily	6.0	6.3	1
PF01368	DHH family	2.5	6.8	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00369	1/2	1	19	1	21	36.3	2.1e-11
PF00095	1/1	87	111	1	40	8.7	1.1
PF01841	1/1	120	130	1	11	6.0	6.3
PF01368	1/1	221	237	1	19	2.5	6.8
CE00368	1/1	54	299	85	333	124.1	6.5e-36
CE00370	1/1	20	338	46	372	336.8	1.9e-109
CE00207	1/1	20	345	47	393	366.8	5.9e-111
CE00369	2/2	20	351	48	379	1143.5	0
PF00864	1/1	20	354	34	395	870.0	7.4e-258

FIGURE 2, page 2 of 2

```

1 TCTCCAAGTC CATGGGTGCC TGGTAGGAGA CAGGGGGATG AATGTGAACC
51 CCTGCATGGC TATAGCCACC TGCCTCCTCC CCTGCCCTGC ATCACTACCT
101 GGCCTATTTT TTGCCCTCTAG AAGCACTGCT TCCTATGCTC CTTAGGACCA
151 CTGCCCCGAT ATGACAGATA AGAACATCGA GGCTAAGGCA ACGCAAATCT
201 TTTCTTTAAA GTCATACAGC TGTCAAAAGA AAGCTGGACA ACCTGGGCAA
251 CATAGCGAGA TAAAAAATTA TTTAAATTAG CCAGATGTGG TAGCCCCCTG
301 TAGTCTCAGC GACTCAGGAG GCTGAGGCAG GAGGCTCACC AGAGTGCAGA
351 GTTCAAGGAT GCAGTGAGCT ATGATCCTGC CACTGCCTGC AAAGCTGGGT
401 GACAGAGCAA GACCTGGCT CTAATAAATG AATACATAAA GTCTCACAGC
451 TAGTGGTAGC TAATCCTGCC AGAGTCAGGC CTCTACCTGT CTGATGACAA
501 ATGGCACACT ATGTCTTTTA ACCTGATTGC AGACCACAAA TGTTTTGTGA
551 ATATTTTCCC CAGGGAAAAA ACCGGAAGTA GTTCTAAATT CTATACATCC
601 ATTATATTAG TTTTACCTGT GGATTGGGAA AACCCAGCTC TGATTGCATT
651 TCAGGGCGGG AGCAGCCTTTG GTGCACTGTC TGGCGGGATT TTCCATTTTA
701 ACCTCCTTCT AGAAGCGCCT TCTCATGTGA AAGTTCTGA TGCCGCCAGG
751 AGCGCCGAGG AGAGGGCAGG GGGCTGGAGA CGCCCCGCAG AGGGCTACGT
801 GCCCTGTGCG ACAGAGGTCT CCTGCCTCCT CGGCGGCGCC AGCCCACTC
851 CCACAACCCC TCGGGGAGAA GCCCCCAAGG GGAGGAGACG GGCCTGGCCC
901 CTGCCCCGAG CACCTTCCGT CTCTAGGTCG GAGTCTGAAT CGGCCTTGGG
951 ACCTGTCTTG GCTTCGGGGA CCCCTGCAAG ACCTCCACAG GCGCGCGTCG
1001 CCTCTTCTCT CTGCTTTTTA TCCTCCCCAG ACCTCTGGCA GGAACCGCTC
1051 ATCGTTACGC CCTTTTCGCA GCCTCAGACC CTGAGGCGGA GACCGCTTGG
1101 CGCCTCACTT AGAGCGCGAC CCGGGGATGT GGGCGGAGTC TGCGGCTGCG
1151 CTGACCAATC GAGTGTGGCG TCCATCGACT GCGCTCTGCC ACGGCAATTA
1201 GCGACGCGCT CCCC CGCGG GGTGCGCCCG GCAACCCAGT GCTGTAGGTT
1251 GCGGTGAGAA CCGTGGCTCT CCTGCGCTGA GGCTCCTCGC CTGAGAGGAT
1301 AAAGTGCACG CGCCACGGGC TATGCACTGG GCTGGGCGCC TTGTGGGCAT
1351 CCTCCTGCC TCCCTAGGGG GTTCCAGCAT CGCCCCCTT TCCTGGACTG
1401 GGAACACGCG CTGACTCCAG GACTTGTGTT GTCTCACTG CACTGGGGAA
1451 GGTGGCGGGG GCAGCTTTTC AGGAGGGCCT GGGGAACCTC GCAGAGCCAG
1501 GTCACCCTCT CACTCTGTGC CTCTTAGTTA TCTTGATGC TCTGGTCTTT
1551 GCATACGCTG CTCCCTGCAC CAGGAACCTC CATCCCCATC TTTGTCTGCT
1601 TGTGGAACCT CAGAAATCTG CAAGGGTCAG CTAGAGGTC ACTTCTTCCG
1651 GAAGCTTTCC TCAACACCTT CCCCCTCTG CTGCTGCTGC CCTCAGGCC
1701 TCCTCTCACA GCACTGATAA CAGCTGTCCG TCTCCACCTT CCCACCACCT
1751 CCACTCCAC CCCAGGAAGT GAGGCCAGAG GGCAGGACA GAGCTGCTGC
1801 TGTCTCTGTG GTGCCAGGGC CCAGCAAAGG GAATGTAGGG AGGTGGGAG
1851 GTGACGGGCA GCTGGGATTA GGGGTTGAGG GCTGGGTGTT GGAGGCTGGA
1901 TCTGGATCCT GCTTTAGTGG AAGTGTCCCT TTAACAGCAA CTGGCCTGGC
1951 CTGGCTCGGG CCTGTCTTG CCTCTGTTC AGCTGCGGCT GCAGCTGCCA
2001 TGTGTACTCA TGTGCCGCA GCTAGCAGGA GCTGGCAGCA TGGGCTCCCC
2051 AGGGGCTACG ACAGGCTGGG GGCTTCTGGA TTATAAGACG GAGAAGTATG
2101 TGATGACCAG GAACTGGCGG GTGGGCGCCC TGAGAGGCT GCTGCAGTTT
2151 GGGATCGTGG TCTATGTGGT AGGGTAAGAG AGAAGAGCTT TTGGCCAGGC
2201 TGGAGGGGCA AGGGAAGAGG TGGGGGTGG GGCTTGTTC TGCTGGGTTG
2251 AAGTTGAGGG TTGGGCTGTT TAGGGGCTGG AGTGAAGGG GGCAGATTGG
2301 GACGGGGTTG GGGAGAGCTA GGCGATACAA GACAGGAGAG CAAGAACAAG
2351 CTGTGTGTTT GTCTGTGTG TCCACTTGCC TCCTTCCAG GCCCCACCC
2401 AGGCCCCACC CAGGGGGCAC ATGACATAGT CCTTAACATC TGTGAGAGCT
2451 GGAGCACTAG GCCCCAGAG AGACCACCAG CTGTATCTCG GGTGAGGAGA
2501 GTCTGTAAGG GGAAGCTGG ATCTAGTCAG GCTGGGGGTG GGTGCTGGCT
2551 AGTGAAGGTG ATTGTCTGAG GGCATTGGCT CTCTGATGCA TGGCTGGAGC
2601 TTCTGTCTCA TTCAGGGGT CTGGAGTGGG AAGTGGGGCC AGAGAGGAGG
2651 TGGGGCCTTC GATGTTGGGC CGGGAGCCTG TAGGGTGTGG GGGGAGAACT
2701 GAGCATGTAG GGCTCAGCTC CGCCCTGTC ACTACAGCTT GGGGACACAC
2751 CACACTGCCC GACTTCTCCT CCCAGGTGG GCTCTCTCG CCAAAAAAGG
2801 CTACCAGGAG CGGGACCTGG AACCCAGTT TTCCATCATC ACCAACTCA
2851 AAGGGGTTTC CGTCACTCAG ATCAAGGAGC TTGGAACCG GCTGTGGGAT
2901 GTGGCCGACT TCGTGAAGCC ACCTCAGGTG GGGGCCCTGA TGTTGCTGAC
2951 GGGGGCGCAA GTCCTTTCCC CACTGACAGC CTGAACACCC GCCATGCAGC
3001 CAGTGTGTGC GAGAGAGAAAG CATGTGATGC CAGAGACGGC TGCGGTTTCT
3051 CAGGAAGGGC TTCACAGAGG AGTGGCACCT GGACAGGACT TTCAGGGATG
3101 TGTAGGAGGT TTTGGGGTGG AAAAAGGGGC CACTCAAGAA GCCAGGCCAG
3151 GGTGGACGCT GCTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCGA
3201 GGCAGGTGGA TCACGAGATT GAGAGTATCC TGGCTAACAC GGTGAAACCC
3251 CATCTCTATT AAAAATACAA AAAATTAGCC GGCATGCTG GTGGGCGCCT
3301 GTAGTCCCAG CTACTCGGGA GGCTGGGGCA GGAGAAATGGC ATGAACCCGG
3351 GAGGTGGAGC TTGCAGTGAG CCGAGATTGC ACCACTGCAC TCCAGCCTGG
3401 GTGGCAAAGC GAGACTCTGT CTCAAAAAAA AAAAAAAGG GCCAGGCCAG
3451 AGAAACTGCA TTTCCAAAGA CTGCCAACAG AAAAGAAGGG AGTGTCCAGG
3501 ACTAATGGCT TGAGCTTGAG AGTGGTGTGA GGTGCTGGG CATGGAACCTT

```

FIGURE 3, page 1 of 10

3551 CCTGTAGCC CTGCTCCCTG ACCTGGGGCA CTACGGTCAG GTGCTGCTCC
3601 TCCCTCTTC TCGGCTGCGT TTTCTCTCTC CCTCCACCCA GCTCATCCCC
3651 AGCCTCAACT GCCACTTCTG CTCCTCTGAT GCCCAGGGTG TATTCACAGT
3701 GATCACCTGC CCAGAGCACA GCTGTCTTCT AGGTGCACAC CCACATGTCC
3751 AAAGATCAAT TATTTTCTCT TCCTGGCATG GCCTCTGTGA CGCCCACTAG
3801 TCATGGTGGC TGTGACATCC ACTAGTGCCT CAGCCAGACC CGTGACTCAC
3851 CCTGGACCCC TTCCTGTCCC TTCCAAGATT TTTCACCACT ACCCATGCCA
3901 TGCCATGCAT GAGACTATGG CCTCCTAGAG GGTCCCTAGA TCCCCCTCTC
3951 GCCTCTCCCC TTAGTGCTCG GTGCACACCA CGCAGCAGCC AAGCTGAACT
4001 TTCACACCAG GCATCATGAG AGCCTGCAGC GCCTGCTTCT ACCCTCAGGA
4051 ATTCCCCCAA CCCTGCCCAT GACGGTGTCC ACACCTTCTT CCAATCCTA
4101 ATGGCTGCCA CTCCAGCAC CATCTGGCCA GCCCTCACCT TCCCTTCTG
4151 GGCATACATT CCCCAAATTC ACAGTGCTCT CACGAGCAGC ACTGGAGGGT
4201 CAGCCTTTCT TTCCAATGTC CTCGGCCACC CGTTGACCAC AGACACAGCT
4251 TTCCCTCTTC TCCCTTGGCC CCTGCCATGC CAGTGTCTGT GTGTGTGAGA
4301 TGGGAGACTC ACCTCGTCTC CATCCTGAGC AGGTGCTGGG CCCAGCTCTC
4351 CCTTGGATCT TCAGTACTAG AAGCAGCAGG CTGTTGGAAT ATTCTGGTTG
4401 GAGCCAGGCA TGGTAGCTGG AGCCTGTAGT CCCAGCTACT TGGGAGGCTG
4451 AGGCAGGAGG ACCTCTTGAG TCCAGGAGTT AGAGTTGCA GTGAGCACTG
4501 ATCACAACAC TACACTCCAG CCTGGGTGAC GAAGTGTAAT CCTGTCTCTA
4551 AATACACACA TACACATGCA CACACACACA CAAATTTTGG TTGAGACAAG
4601 AGACTTGTCT CAAGAGATGG ACATGGGCAC AAGGCTTCTT GGTCTCAAAA
4651 ATGGCCAGAA CCACTGCCAG CCTCCCATCT CTGCTTCAGT CTGCCTTACA
4701 GGGGGACAGG GTTAATGACT TGATGGGGCC AACATCCCTT CCCTCATAAA
4751 CCAGGCTGCC GGCTTCCGGC CTTTCCAGTC AACACGAGCC CAGCCAGGCC
4801 AACCTTGAGA CTTGCCTCCT AGGGAGAGAA CGTGTCTTTC TTGGTGACCA
4851 ACTTCCCTGT GACGCCAGCC CAAGTTCAGG GCAGATGCCC AGAGGTGAGT
4901 TTACCAGGGA TCCTCCCAGC GGTCCCTTG TTCCTCCATC AGCCCCAGGT
4951 GGCCACCCGT GTTTCCCTTT CCCCTTCCCA GGTGGCTGAA GGCTCAGCCT
5001 GTGCTCGGTG TCCCCCAGGC ACTGGGCTAC ATCTTTTCTT GAATCATTAT
5051 GTTCAGTCTT CACATATCCC CTGCCTGGTA GGAAGTCCTG TGATCCCCAT
5101 TTCAGAGGAG AAGACTGAGG CTCAGTGAGG TTGAGTCACT TTCTTAAGGC
5151 CTCCAGGCCT GTGGGTGACA GGACCCGAG CTCTGGGCAG CAGCAGTTCC
5201 CATGAGGTGT CAGGCCCTC CCATCCTGGT CCGCTCTCTG GGTACTCTCC
5251 AGGTTGGTAG TGTGACACCC AGAGCTGCGC ACATGCTCAG GGAGGTTCTA
5301 ATAGCAAGAG CCAAGCTGGA ATATCACCTC CCCTTGTCTG TGCCACGCCT
5351 CTATTAATAT GTCCTGAGGC AGCTTTCATC TTTGTGGGCC AACACAGCAC
5401 ACTCTTGCTC ATGGTGAAAT CAGGATTGCT TATGATTCTT GGATAGTTTT
5451 TTTTGTTTTA TTTTGTGAGC GGAGTTTCAC TCTGTACCCC ACGCTGGAGT
5501 GCAGTGGCAG ATATCAGCTC ACTGCAAGCT CTGCCTCTCA GGTTCACGCC
5551 ATTCTCTGCT CTCAGCCTCC GGAGTAGCTG TACTACAGG CGCCTGCCAC
5601 CAGCCCGAGT TAATTTTTTT TTTTTTTTGT ATTTTTAGTG GAGACGGGGT
5651 TTCACGGTGT TAGCCAGGAT GGTCTCCATC TCCTGACCTC ATGATCCACC
5701 TGCCCTCGGC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCACGCCC
5751 GGCCTGATTT CTGGATAGTT TTTACATCAA CCGTGGTCAA GCCAGAGTCC
5801 CCCACCTTGT TCTTCTTCAT TTCTGATCCA GAAATGCTGA TTCTCCCCCT
5851 GACATTTTAC CTTTTCCCCT TGCCTGGGGA TGTCCCTGGG ATCCTGCATC
5901 TGTACAGAG CATGCTCATT CTCTCCAGCT GTGAATTTTG TTTGAACTAT
5951 TGGGACTCAG GACATAGTCC TGAAAGTTTA CCTCCACAGT GACATCTTTA
6001 GGCAAGTCCA ACATTTACGT GCCTCCTGGG CTGGAGGGTC GTTGTGAGA
6051 CAGCTGTCCC CTGAGCCCTG GTGGCTGGTC CTAGCACAGT TGCTGGAGAC
6101 ATCCCATGTC CGTAGTTGGA AATATGCACA AAGGATTGCT TACTCTTTTT
6151 GTTTGTTTGT TTTTTTGAGA TGGAGTCTTG CTCTTGTCCC CAAGGCTGGA
6201 GTTCAATGGC ACGATCTCGG CTCACTGCAA CCTCCGCCTC CTGGGTTCAA
6251 GCAGTTCTCC TGCTCACCCC CTGAGTAGCT GGGATTACAG GTGCCCGCCA
6301 CTGTGCCCAG CTAATTTTGT TATTTTAAAGT AGAGACGGGG TTTCACCATG
6351 TTGGCCAGGC TGGTCTCGAA CTCCTGGCCT CAGGTGACCC ACCAGCCTCG
6401 GCCTCTCAAA GTGCTGGGAT TACAGGCGTG AGCCTGCCGA GAGCTTGGTC
6451 GGGGAGACCT GAACCCAGCG GTGCTAAAGG AATTAAAGAC AAACACACAT
6501 AAATATAGAG GTGTGGAGTG GGAAATCAGG GGTATCACAG CCTTCAGAGC
6551 TGACAGCCTC GAACAGATTT ACCACATAT TTATTGACAG CAAGCCAGTG
6601 ATAAGCATTG TTTCTACCAG ATTATAGATT AACTAAAAGT ATTCTTTATG
6651 GGAACAAAG GGTAGGGCTC TGGTTGGTTA TCTGCAGCAG GAGCATGTCC
6701 TTAAATCACA GATCGCTCAT GCTATTGTTT GTGGTTTAAAG AACGCCTTTA
6751 AGCGGTTTTT CGCCCTGGGT GGGCCAGGTT TTCCTTGCCC TCATTCGGGT
6801 AAACCCACAA ACTTCCAGTG TGGGTGTCGT GGCTATCACA AACATGTCAC
6851 AGTGCTGCAG AGATTTTGT TATGGCCAGA TTTTGGGGGC CTCTTCCCAA
6901 CATGAGCCAC TGTGCCTGGC AGGATGTGCT TACTCTTGGT GAACCCACAC
6951 AATGCTCTTC TCTTCTTAA TGCTCAGATG TGCAATTTAG GTTCAGTTTG
7001 TAGACCGTTC TGAATTTTGG CTGGATCTGT GGGTCTGTGT TTTTCAGAAT
7051 CTGTGCAATT CCTCTTTGTC TGCAACCACA CTTCTGGCTC TTCCCATGAA

FIGURE 3, page 2 of 10

```

7101 ACGTCAGGGC TGGGTCGTAA TTATCAGATC TGACAACCTG GCTTTCCCGG
7151 AAGACCAGAG TTCTGCCAGC TCCTCTAGGG ATCCTGGTGC CTGATCCCTC
7201 CCTTACATGC ACCATGCTCT TTATAGTGTC ACCTCCCTCA GCAGACACCG
7251 CTGAGCCTCC CCGCTGGGCC AGGGGGCTAG CTAGGCTAAA TTCACAAAAC
7301 TCCATCTCCC ATACTTCAAA GACCACCCAC ATGGACAGCC CAGCCCAGGT
7351 GGCAGGTCGG ATGATGGGAC AGAGGCTGTA GGTGGGGGAC CTAGGGCTGC
7401 ACTTGAGCAG AATCTTTTTT TTTTTTTTCT TTTTTTTTTT TTTGAGACAG
7451 AGTCTCGCTC TGTACCCAGG GCTGGAGTGC AGTGGCGTGA TCTCGGCTCA
7501 CTGCACACCT CCACCTCCTT GGTTCAGCG ATTCTCCTGC CTCAGCCTCC
7551 CAAGTAGGTG GGACTACAGG CACACACCAC CACACTCGGC TAATTTTTGT
7601 ATTTTAAATA GAGACAGGGT TTTGCTGTGT CGGCCAGGCT GGTCTCGAAC
7651 TCCTGACCTC AGGTAATCCG CCCACCTTGG CTCTCAAAG GTTTGGGATT
7701 ACAGGTGTGC CAGGCCAAGC AGAATCTTAA AAAAAGGTGG GGAGAAGCTG
7751 GTGACAGGTT GATTTGGTT GAAGCAGGAT GTCGACACAG AGGGGGCTTG
7801 GTGGGTAAAG GCCCTGAGCT GTGTGAGGTG AGGTGCCTTT AGGGCTACCT
7851 GCACTGGGTG GGAGCTGAAG TGAAGATTG GACTGGGGTG GGAAGAAGGT
7901 AGTTCAGGAT TTCAGGGGCC CCTGTAAGCC CCACTAAGGA GCTAAACTGT
7951 TTTTGTFTGT TGTFTTCTT TTTCTCTTTT CTTTTTTTTC CTGTAGCAAT
8001 GAGGTCTTGC TTTGTTGCCC AGGCTGGTCT CGAACTCCTG AGCTCAGGCA
8051 ATCCGCCTAC TTGGACTCT CAAAGTGCTA GGATTACAGG CGTGAGCCAC
8101 TGTGCCTGGC AGGAGCTAAA CTTGATTAGA GGAACAGAAG AGAGCCACAC
8151 GTGGGTCTAG AGGCAGGGTG CTCAGTTTCC TGCACATTGG GATGCACCAC
8201 TTGGGCTGCT GGGCATAGGT GGATGAGGTG ATGGGAAGAC GTGGGGGCCC
8251 CACTGGTGGT CACTGTGGGG TCTAGTTGGA GGAGACGGTA GCCCAGCTGG
8301 GGTGAAGAGG AGAGGCAGAC ACAGGACATA GGTAGGGACA AAGAAGCAGA
8351 GCATGTGGCT CTGCTCCGAC CTCACCCCAA TCACGACGGC CCTGTCTTTC
8401 AGAAAGTCCC ACCGCCTCAT TCTGGCTTCT CAGAGGCCCT CAGCCTTCCT
8451 TGGCCCCCTG GTGCTGGTGT TCTTCCTGCT GCCCCTGAGC TGAGTGCCCT
8501 GGGCAGCAGT GTCCATCCTC AGTTGGGGCA GGACCATGCC TGGGAGAGTG
8551 CCCGATGCTC AAGGGTGCC TCGTCTCTGG GGTCTGGGAC CCCAGAAAGC
8601 TCACCTGTCC TCCCCTTCTG CCAGAGCCCC ATAGTCCCAT GCCTCTGTGC
8651 AGGCATTAAAT GTCCCCAGGT TACAGAAGAG CGAGCAGGAA GGAGTAGCCT
8701 GTGGTCCCTC AGCAAGGGTG TGGGGTCTCT CTCAATACC CAAGCCCCTG
8751 ACTCTAGGGC CTGATCTTT GTCACTATG TCCCCATGCC GGGCATCAAA
8801 AACTCACCTC CCCAAGGTAT CTTACCTTTC CCTGATCTGT CATCCAAATT
8851 GGACCAGAGG AGCTAGACCT GGAAGAATCA CTTCGCGATC CACCAGGGAC
8901 AGAACTGTCA GAGGGGAAGG GGCAGGGTGC GTGTCTCAC GCCTGTAATC
8951 CCAGCACTCT GGGAGGCTGA GACAGAAGGA TTGCTTGAGG CCAGGAGTTA
9001 AAAACCAGCC TGGTCAACAT AGCAAGACTC CATCTCTACA AAAAAAAAT
9051 ATTAAAAAAT CAGCCAGGCA CAGTGGTGTG TGTCTGTAGT CCCAGCTACT
9101 GGAATACTG AGGTGAGAGG ATTGCTTAAG CCGGGGAGGG CGAGGCTGTA
9151 GTGAGCCATG ATCATAACAC TGCACTAGAG CCTGGACAAC AGAGTGAGAC
9201 CGAATCACTA AAAATAAATT TTTTGAAAAA GGAGGAAAGG GGTCTCCCTT
9251 TGTCTTTGAA ATACAGTACT GTACCTTCAT CTGGCCAGGG CATTGCTCCG
9301 CTCCTCCTC TGACCACCTC CTTTTATTG CACCCTCCAG CTTTCTGTG
9351 TGGCCCCACA CTCAGGGTAC TCTGGCGGCG GGGTGGTGAG GTTGTTTAAG
9401 GTGGGAAGGG GGCCTGTCT TCCACCTTG AACCTCCCTG CCTTGAGAC
9451 TGGGCTGTGG AGGGGAGACA TCCCCTGTGC CATTGGTGAC TGCTCTCTCT
9501 CCCACCTCAG CACCCGTCCG TCCCACTGGC TAACTGCTGG GTCGACGAGG
9551 ACTGCCCCGA AGGGGAGGGA GGCACACACA GCCACGTAA CTGTGGGCTC
9601 TGTCTTCCAG TGCCCCTAGC AGGGTGGGG CCGGGCTGGG ATCCTGGGTG
9651 GCTCCTGAGT GCAGGCCCTG CTCGCCTCTG TCCCTGCATC TCTCTTCTG
9701 CCAACAACCC CTGGCTGAA GGCTTCCCA GGCTTGAGA GATTGAAAG
9751 TCTGGAGTTC ATCTTTTGT TTTCTAGGTG AAAAACAGGC CAGTGTGTGG
9801 TGTCAATGG GACCCACAGG ACCTGTGAGA TCTGGAGTTG GTGCCCCGTG
9851 GAGAGTGGCG TTTGTGCCCTC GTAAGTGTCC CCACAATCCC CTACCCCAAC
9901 TGGCGCAGGG CCCCAGGCCT GGCAGAGGCT GTCACCTCCC TTCCACCTGC
9951 AGGAGGCCCC TGCTGGCCCA GGCCAGAAC TTCACACTGT TCATCAAAAA
10001 CACAGTCACC TTCAGCAAGT TCAACTTCTC TAAGTAAGCA GAGTGGGTCT
10051 CATCTGCCCC AAGACCCTCC TTGTCCCCTA CCTCATCTGA CCTTTCCAC
10101 TCTTCCCAGG TCCAATGCCT TGGAGACCTG GGACCCACCC TATTTTAAGC
10151 ACTGCCGCTA TGAACCACAA TTCAGCCCCT ACTGTCCCGT GTTCCGCATT
10201 GGGGACCTCG TGGCCAAGGC TGGAGGGACC TTCGAGGACC TGGCGTTGCT
10251 GGTGGGTCCC AAGTTGGGG CAGGGTTCCT AGAGGGCTCT GGGAGAGGGT
10301 CCCGGGCCCA CACACCGGTG GAAAAGCTAT GTGCTATGTG CAGGGTGGCT
10351 CTGTAGGCAT CAGAGTTCAC TGGGATTGTG ACCTGGACAC CGGGGACTCT
10401 GGTGCTGCTC CTCACTACTC CTTCAGCTG CAGGAGAAGA GCTACAACCT
10451 CAGGTGAGGC CCCACTGCTC CCAGTGCCCA GCTGCTGGGC CCATCGCCCT
10501 CTCACTGTGG CGCCAGGAC AGACCACACC CAGGCCAGG CCTCTAGATA
10551 TTCCACTACG TGTGCAAGGG GGTCCAGGA GCAGGAGAGA GCTGTTCTCA
10601 ACCCCACATC CTCAGCACA GGCTCCGTCC TGCTGCCCA AGTCTGAGC

```

FIGURE 3, page 3 of 10

```

10651 CCTCCACCCC ATCTGTCCCA GGCCCTTGCC CAGCTCAGGC TCCTCACTGC
10701 CAGCCCTTCC TCCACCCAC CTCGCTTCTA GTATCTCCCC TCCACAGCAA
10751 TGGGGTGTTC CATTCTTACT TTCCCTTCTT CCCCTTCAGC TTTGTTTTTT
10801 TTTTTTTAAG ACAGAACTCTC ATTCTGTACAC CCAGGCTGGA GTGCAGTGGC
10851 CCGACCTCGG CTCACGTGTA CCTCTGCTTC CTGGGTTCAA CCGATTCTCC
10901 TTCTCTAGCC TACTGAGTAG CTGGAATTAC AGGTGCTCGC CACTACTCCC
10951 AGCTAATTTT TATATTTTGG TAGATAGAGA TGGGTTTTCA CAATGTTGGC
11001 CAGGCTGGTC TCAAACCCCT GACCTCAGGT GATCCACCCA CCTCAGCCTC
11051 CCGAAGGGCT AGGATTACAG ACGTAAACCA CCATGTCTGG CCTCCCTTCC
11101 GCTTTTACCT AAACCTTTTT TTTTTTTTTG AGATGGAGTC TCACTCTGTC
11151 GCCCAGGCTG GAGTACAGTG GCGGGATCTC AGCTCACTGC AAGTTCCGCT
11201 TCCCGTGTTC ACGCCATTCT CCTGCCTCAG CTCCCAAGT AGCTGGGACT
11251 ACGGGTGACG GCCTCCACGC CCGGCTAATT TTTGCATTTT TAGTAGAGAC
11301 AGGGTTTAC CATGTTGGCC AGGATGGTCT CGATCTCTTG ACCTCGTGAT
11351 CCACCTGCCT CAGCCTCCCA TAGTGCTGGG ATTACAGGCG TGAGCCACCA
11401 CGCCCGACCT TTTTTTTTGA AACGGAGTTT TCACTTTCTT GTAGTCCAGG
11451 CTGGAATGCA ATGGCGTGGT CTGGGCTCAC TGCAACCTCT GCCTCCTGGG
11501 TTCAGGTGAT TTCCAGCCT CTGCCTCCAG AGTAGCTGGG ATGACAGGTG
11551 TGCACCACCA CACCCAACTA ATTTTGTAT TTTTAGTAGA GATGGTGTTC
11601 TGCCATGTTC GCCAGGCTGG TCTCGAAGCT CTGACCTCAG GTGATCTGCC
11651 CACTTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACCAAGCCT
11701 GTTTTTTTTG TTTTTTTTTT TTTTTTTTTT TTAGATGAAG TTTTGCTCTT
11751 GTTGCCCGA CTGGAGTGCA GTGGCCCGAT CTCGGCTCAC TGCAATCTTT
11801 GCCTCTCGGG TTCAAGCAAT TCTCCTGCCT CAGCCTCCTG AGTAGCTGTG
11851 ATTACAGGTG CACACCACCA CACCCAGCTA ATTTTGTGT TTTTACTAGA
11901 GATGGGGTTT CATCATATTG GTCAGGCTGG TCTCGAAGTC CTGACCTCAG
11951 GTGATCCACC TGCCCTCAGCC TCCCAAAGTG CTGGGATTAC AGGTGTGAGC
12001 CACTGTGCTT GGCCTCAAGT TTCATAAATT GCATTTATTA TCATGTCTTT
12051 GAGTCTTCTA AGCAGATCTA TTGGATCCTT CTGCCACCGA GCGTCACCTC
12101 GTCATGCGAG CAGGCACACA CGACCACAG GCCTGGGGAT GATGCCCTC
12151 AACATAGCTC ACTGCACCCC GTCTGATCTG GCTTCCCAA CCTCCCCAGC
12201 CCTTCGAAAC CACGTGGGGC TGGTCCAC CCACATCCTG TTCCCTGAC
12251 CTCTGTGCTG GCAAACCACC TGTGTGCATG TTCCTTCAGG CCCAGCCTCA
12301 TGTCCCTTCC AGGAAGTCTA CCCCAGTTCC CAGGGAAGAG TGAGTTCCCA
12351 TCTCTGGAAT CCCTCAGCCC TGAGCCTGCC CCTTCACATC CCCCCTGCT
12401 GGGTCTGTTC AGGGATCCT CTGTCCCGG TCCTCTCAGC AGGCAGGGAA
12451 CTCTTGAGG ACAGTCTTC GTTTGCTTTT TCTGTTTCT CACCAATTAC
12501 ATAGGGCTGA GACCCAGGAC TCAGGCTTGG GCTGGGGTT TATAGAGTCA
12551 ATTGACAAGT TGGACAGAGG TCTGGCAGGG CCAGCCAC CTGGGGGTGG
12601 GCAAAGCAGG TCACCAGAGC CTCTTTTCTT GCCCACAGGA CAGCCACTCA
12651 CTGTTGGGAG CAACCGGGTG TGGAGGCCCG CACCCTGCTC AAGCTCTATG
12701 GAATCCGCTT CGACATCCTC GTCACCGGGC AGGTAGGCAC AGGTAGGGGT
12751 CAGGCCGGGG ATGGGATGGG GCAGGCAGAC AGGCTGGAG GAGGCATGAG
12801 GCTGACAGTC GTGGGCTGAG AGGTTCAAGT CAGATCTCTC TCAGGCAGGG
12851 AAGTTCGGGC TCATCCAC GGCCTCACA CTGGGCACCG GGGCAGCTTG
12901 GCTGGGCGTG GTGAGTGCGA GCACTGTGGG CACCTGCAGG CTGCAGTGAG
12951 TGCTGCTGAC CAGGGTGTGT CCAATGCATG CTGGAGCCTC CGGTGCCTGC
13001 ACATTGAGTC TCGGGGTGCA GGCTGGGGAG GTGGCAGGAG AGCAGGCTCG
13051 GGGCTGGGAA CATGGGTTGG CCTGCCTCT CCCAGGTCAC CTTTTCTGT
13101 GACCTGCTAC TGCTGTATGT GGATAGAGAA GCCCATTCT ACTGGAGGAC
13151 AAAGTATGAG GAGGTGAGCT GAGGTCGCTC TGCTTGACC CTGGGTCTG
13201 CCACACTTAG GAAGATGTTG GCTGGATCCC TGACCTGCTG TCCTCATCTG
13251 CAGGCCAAGG CCCCAGAAAGC AACCAGCAAC TCTGTGTGGA GGGAGCTGGC
13301 CCTTGCATCC CAAGCCCGAC TGGCCGAGTG CCTCAGACGG AGCTCAGCAC
13351 CTGCACCCAC GGCCTGCTG GCTGGGAGTC AGACACAGAC ACCAGGATGG
13401 CCCTGTCCAA GTTCTGACAC CCACTTGCCA ACCCATTCG GGAGCCTGTA
13451 GCCGTTCCCT GCTGGTTGAG AGTTGGGGGC TGGGAAGGGC GGGGCCCTGC
13501 CTGGGGATCT CAAGGATGAG GCCCCAGCAT GGAGGATTGG GGGTAGAATT
13551 CCACCTTGA ACCCCAGCAG ACAGTCCCTC CCTGACTCC CACCTTGGTA
13601 GGGTGCTGCC TCAGGGAGCC ATAGAAGTCG GCTGTGTTTT GAGACGGCGA
13651 CAGAACCCTA CCGTGGAGA CTGGGAGAGC CCAGCAGGCA CCTGTATTGC
13701 AGGGCTCCGA CTGCATGTGG CAGGGGCTCC TGCTGCTCT GGGCCTGGAG
13751 GTCTCTCTCC CAGTGCTCTG TCCCAGTGT TCCTAGCAGA GGTATGCTTA
13801 CCAGCTGTCA GCACAGACCC TCCTGCTGCC TGGGTCCTGG CCCTCCTCCC
13851 CCATCTGCAC CCCCATCATA GGTAGAGACC CACCCTCCC ATCGGTCTTA
13901 CATGGGGCTG TGCAGCTGGA GCCAAAAGG CAAGGTAGAA AGAGGAGTGA
13951 TGGGGGAGGG GGATTGTTTC AGCTTCTCTG GTGCTGTGAT GCCCCAGGAG
14001 AGTCCTAATC TAGGGAATGG GGTGGAGTAG GCAGATAATC CACCTCCCTA
14051 TCCCCCAGC AAGGGCGGAG CATGTGCTTT GGGCCACAC CTGCTTAGTT
14101 TATGAGGACC GGCTGCTTTC CAGTGCTAGC CCTTTTGCCA TGGAGTCTG
14151 GGAGAGAGAG CAGAGGGCGG CAGGGCTAAG TTGGTGATCA TTGGGTCTT

```

FIGURE 3, page 4 of 10

14201 CAGGACCTTC TATATCCCTC CTCGGTAACC CCCCAGCCCA ACCCCTTGGA
14251 ATCTTTCTCT CAGGCTTCCT GAGAGCCCTG GGGGTGGGAG GCTGTGGGAG
14301 GCTGTACATC TGAAATTCAC TTCAGTCCAA GTCATACCTA GGAAGCTGTC
14351 TGGGCAGGTG CTCGAGGGAG GCCCTGGCTC TGATCCCAAG CTGGATGGAG
14401 TGGCTGGAAG GAATGGTTCC AAACAACACC ACCGAGATCT CCCTCAGGCT
14451 GGGCAGGTTT TGCAGCTGGA ATTCTCCTCT TGGTCCCAGG GCGGGGAGG
14501 GAATTCTAAG TGTCCACCCC AGGAGGGCAA GGGGCTGCTT TCCACTGTGG
14551 GTACCTGTGT ATCAGGGCAA GCTGTGGAGG GCCAGGGGTG GGGCTGAGAC
14601 TGGGCTGACA TCTAGAATCA CCTGCCACCT GGAGCCTCAG TAAAATGCCT
14651 GGGGTCCCTG CTGCCTCTCA ATCTCCAGAG CCATGTCCAT GGGGAGGTGG
14701 GCTCTGAAGG GCGAAGGTGG GAGAGCAGGG CCCCTGAGGC CTGGGTATCC
14751 AAGGAGGGGC ACGTGACCTT GATTCTCCTT GGGGCCCAGA GGAAGCTGAT
14801 GTCATGGCTG GACAAAGTCA CCGAGTAAAG CCAGCAAAGC CACCCTCTTC
14851 CTGTGTAGTC TGTACAGGCA TGAAGGAAAG GTGGGGGGC ATCTATGGTA
14901 GACATGGCAC AGCCATGAAG AGACCACTGG GGTGGTGCAG GGTGGACTTG
14951 GGGACCCCTAC CCCTGAAGAC TGAGGCCCTG CAGCTACCAG GTGGGCTAGA
15001 AGGTAACCTG AACAGGCCCTG GGCACCTTGT CACCCATGTA GGAGCATGAG
15051 GGGCAGGTTT TTTTCACTC AAAGCCCTTG AAGAGTGGGC AAAGACAGCA
15101 AGAGAGCTGC AGCCTGGGCC CGAGCTCAGA AACAGCTGTC GCCTCAGTCT
15151 GCGCACAGGC ATGCACCCCA GGGTAGTGCC TGCAGGGATG CATGTGTCCC
15201 CGTGGGGGTG CCTGTGCCAG GCAGGCCCTCA GGTGCATGCC ATGCTCAGAA
15251 CCTGTCTGCC CTTCCTAGGC AGCCTCCTTG GGGCCCAAGC TCTGCTCCCT
15301 GGATCTGCCA CCTAGCAGAC GTGGGGAGCC TGACCCCATG CCTGTCTATG
15351 AACCTCTCCT GCTGTGTGTG TGTGGCTCCC CTCTTCACTG GGCACCTGGA
15401 TCCAGGCCCA CCTGTGTCCC TGAAGCAGGG TGGTCCCAGG ACTGGCACCT
15451 ACTCTTTTGA GAGCCCCAGC ATCTTTGATG TGGATTGGAG ACAATTGCCT
15501 GGTTCCTCTG GGCAGGTGAA GACTTGGTGC CACAAAGAAT GCCACAGTGG
15551 ATACGCCAGC AGGCCACATG GCTGGCCAAG CAATTATTAT TATGGATCCC
15601 TTGGGCTGTG GGCCTTCCCA TCCACCCAC CACAAGTCCC CAGGTAGCTG
15651 GAGCTGATCA TAAACAAGAA GGCTCTGGGC AGAGTCCATG GCACCAAGCA
15701 CAGCAAGGC CCACTCCTGA AGACCCGAG CCCAGCCCTT GGATGAAGGT
15751 CCTAAGGTCC TGAGGACTCC CCAGCCTGTG CAGGCCCTGA AACCCAGGCT
15801 GCCCACAACA GAAGGGGCTC TCGGCTTGTG TGGCCTCTCT GGCCTCCCAA
15851 GCAGGTGTGG GAGGGCGGGG CAAGTGTGGG CTGATCAGT ACTCCATATG
15901 GCCAGGGTCC TGTGCTGGTG CCTGGCTGGG GGGCTGCATA GCCTGCACTG
15951 TCTCTCCAG GCTGCCCTG GGGAAATACCA CGTAGTGTGT GGAGTTCAGC
16001 CCTGGCAGCT CCCGCTGGTT CTCCTTGCTA TGCCGGATGC CATAGCCGAA
16051 ATACACTGCA AGTCCTAGAC AGGGCAGGAG GCAGGGCATG AGCCTGAGGT
16101 ACAGGTTCCA GCCCTTCTG TCCTCTTTCG CCTCCTCCTG ACCCGGTCC
16151 CAGCCTGGCC CCCACTCACC CATCAGCAGC CAGATGGAGA AGCGCACCCA
16201 GGTGAGATAG CTAAGTTTCA GCATGAGGCA GATGTTGAGG ACGATGCTCA
16251 GGGCTGGAAT CAGGGGAACC ATGGGGATCT GAGGAGGAG AGGCAGGGCA
16301 GGGCTGGGCC GGGCTGCAGG AAAGATCTGC CAGCCCAGGG CTCACCTTCT
16351 CGGGAATCCA TAGAGCCTTT GTTCTCTACG GGAGATTGTG GAGACATGTG
16401 CTCACTCACC ATGCAGAAAG GGGTGCAGGA TGGGTGTGTG GTCCTCCCC
(SEQ ID NO: 3)

FEATURES:

Start: 2040
Exon: 2040-2095
Intron: 2096-2776
Exon: 2777-2927
Intron: 2928-4822
Exon: 4823-4894
Intron: 4895-9510
Exon: 9511-9586
Intron: 9587-9776
Exon: 9777-9870
Intron: 9871-9952
Exon: 9953-10033
Intron: 10034-10109
Exon: 10110-10251
Intron: 10252-10343
Exon: 10344-10453
Intron: 10454-12638
Exon: 12639-12732
Intron: 12733-12844
Exon: 12845-12910
Intron: 12911-13085
Exon: 13086-13163
Intron: 13164-13253

Exon: 13254-13448
Stop: 13449

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
136	C	T	Beyond ORF (5')			
253	T	C	Beyond ORF (5')			
573	C	T	Beyond ORF (5')			
2000	A	G	Beyond ORF (5')			
2222	G	C	Intron			
2783	G	T	Exon	21	A	A
3199	G	A	Intron			
3307	C	G	Intron			
5012	C	G	Intron			
6169	G	C	Intron			
7647	A	G	Intron			
8638	C	T	Intron			
9409	T	G	Intron			
10504	A	C	Intron			
10971	T	-	Intron			
12609	G	A	Intron			
13367	T	A	Exon	378	T	T
14191	T	C	Beyond ORF (3')			
14227	A	G	Beyond ORF (3')			
15027	T	C	Beyond ORF (3')			
15441	A	C	Beyond ORF (3')			

Context:

DNA
Position

136	TCTCCAAGTCCATGGGTGCCTGGTAGGAGACAGGGGGATGAATGTGAACCCCTGCATGGC TATAGCCACCTGCCTCCTCCCTGCGCTGCATCACTACCTGGCCTATTTTTCCTCTAG AAGCACTGCTTCCTA [C, T] GCTCCTTAGGACCACTGCCCGCATATGACAGATAAGAACATCGAGGCTAAGGCAACGCAA ATCTTTTCCTTAAAGTCATACAGCTGTCAAAGAAAGCTGGACAACCTGGGCAACATAGC GAGATAAAAAATTATTTAAATTAGCCAGATGTGGTAGCCCCCTGTAGTCTCAGCGACTCA GGAGGCTGAGGCAGGAGGCTCACCAGAGTGCAGAGTTCAAGGATGCAGTGAGCTATGATC CTGCCACTGCACGTGAAGCTGGGTGACAGAGCAAGACCTGGCTCTAATAAATGAATACA
253	TCTCCAAGTCCATGGGTGCCTGGTAGGAGACAGGGGGATGAATGTGAACCCCTGCATGGC TATAGCCACCTGCCTCCTCCCTGCGCTGCATCACTACCTGGCCTATTTTTCCTCTAG AAGCACTGCTTCCTATGCTCCTTAGGACCACTGCCCGCATATGACAGATAAGAACATCGA GGCTAAGGCAACGCAAACTTTTCCTTAAAGTCATACAGCTGTCAAAGAAAGCTGGACA ACCTGGGCAACA [T, C] AGCGAGATAAAAAATTATTTAAATTAGCCAGATGTGGTAGCCCCCTGTAGTCTCAGCGAC TCAGGAGGCTGAGGCAGGAGGCTCACCAGAGTGCAGAGTTCAAGGATGCAGTGAGCTATG ATCCTGCCACTGCACGTGAAAGCTGGGTGACAGAGCAAGACCTGGCTCTAATAAATGAAT ACATAAAGTCTCAGAGCTAGTGGTAGCTAATCCTGCCAGAGTCAGGCCTCTACCTGTCTG ATGACAAATGGCACACTATGTCTTTTAACTGATTGCAGACCACAAATGTTTGTGAATA
573	TAAATTAGCCAGATGTGGTAGCCCCCTGTAGTCTCAGCGACTCAGGAGGCTGAGGCAGGA GGCTCACCCAGAGTGCAGAGTTCAAGGATGCAGTGAGCTATGATCCTGCCACTGCACGTAA AGCTGGGTGACAGAGCAAGACCTGGCTCTAATAAATGAATACATAAAGTCTCAGAGCTA GTGGTAGCTAATCCTGCCAGAGTCAGGCCTCTACCTGTCTGATGACAAATGGCACACTAT GTCTTTTAACTGATTGCAGACCACAAATGTTTGTGAATATTTTCCCAGGGAAAAAAC [C, T] GGAAGTAGTCTAATAATCTATACATCCATTATATTAGTTTTACCTGTGGATTGGGAAAAAC CCAGCTCTGATTGCATTTCAGGGCGGACAGCCTTTGGTGCACTGTCTGGCGGATTTC CATTTTAACTCCTTCTAGAAGCGCCTTCTCATGGTAAAGTTCCTGATGCCGCCAGGAGC GCCGAGGAGAGGGCAGGGGGCTGGAGACGCCCGCAGAGGCTACGTGCCCTGCTGGACA GAGGTCTCCTGCCTCCTCGGGCGCGCCAGCCACCTCCCAACCCCTGCGGGAGAAGCC

FIGURE 3, page 6 of 10

2000 CTCTCTCTCACAGCACTGATAACAGCTGTCCGTCTCCACCCTCCACCACTCCACTCCCA
 CCCCAGGAAGTGAGGCCAGAGGGCAGGGACAGAGCTGCTGCTGTTCTCTGTGTGCCAGGG
 CCCAGCAAAGGGAATGTAGGGAGGGTGGGAGGTGCAGGGCAGCTGGGATTAGGGGTGAG
 GGCTGGGTGTTGGAGGCTGGATCTGGATCCTGCTTTAGTGGAAGTGTCCCTTTAACAGCA
 ACTGGCCTGGCTGGCTCGGGCCCTGCTTTGCCTCCTGTTTCTGCTGCGGCTGCAGCTGCC
 [A, G]
 TGCTGACTCATGTGCCCGCAGCTAGCAGGAGCTGGCAGCATGGGCTCCCCAGGGGGTACG
 ACAGGCTGGGGGCTTCTGGATTATAAGACGGAGAAGTATGTGATGACCAGGAAGTGGCGG
 GTGGGCGCCCTGCAGAGGCTGCTGCAGTTTGGGATCGTGGTCTATGTGGTAGGGTAAGAG
 AGAAGAGCTTTTGGCCAGGCTGGAGGGCAAGGGAAGAGGTGGGGGGTGGGGCTTGGTCC
 TGCTGGGTTGAAGTTAGGGTTGGGCTGTTTAGGGGCTGGAGTGAAGGGGGCAGATTGG

2222 AGTGTCCCTTTAACAGCAACTGGCCTGGCCTGGCTCGGGCCCTGCTTTGCCTCCTGTTC
 GCTGCGGCTGCAGCTGCCATGCTGACTCATGTGCCCGCAGCTAGCAGGAGCTGGCAGCAT
 GGGCTCCCCAGGGGCTACGACAGGCTGGGGGCTTCTGGATTATAAGACGGAGAAGTATGT
 GATGACCAGGAAGTGGCGGTTGGCGCCCTGCAGAGGCTGCTGCAGTTTGGGATCGTGGT
 CTATGTGGTAGGGTAAGAGAGAAGAGCTTTTGGCCAGGCTGGAGGGCAAGGGAAGAGGT
 [G, C]
 GGGGGTGGGGCTTGGTCTGCTGGGTTGAAGTTAGGGTTGGGCTGTTTAGGGGCTGGAG
 TGGAAAGGGGGCAGATTGGGACGGGGTGGGGAGAGCTAGGCGATACAAGACAGGAGAGCA
 AGAACAAGCTGTGTGTTTGTCTGTGTCCACTTGCCCTCCTTCCCAGGCCCCACCCAG
 CCCCACCCAGGGGGCAGATGACATAGTCTTAAACATCTGTGAGAGCTGGAGCACTAGGC
 CCCAGAGAGACCACAGCTGTATCTCGGGTCAGGAGAGTCTGTAAGGGGGAAGCTGGAT

2783 GTATCTCGGGTCAGGAGAGTCTGTAAAGGGGAAGCTGGATCTAGTCAGGCTGGGGGTGGG
 TGCTGGCTAGTGAAGGTGATGTCTGAGGGCATTGGCTCTCTGATGCATGGCTGGAGCTT
 CTGTCTCATTACAGGGGCTGAGAGTGGGAAGTGGGGCCAGAGAGGAGTGGGGCTTCCGA
 TGTGGGCGGGAGCCTGTAGGGTGTGGGGGAGAACTGAGCATGTAGGGCTCAGCTCCG
 CCCCTGTCTACACGCTGGGGACACACCACACTGCCGACTTCTCCTCCCCAGGTGGGC
 [G, T]
 CTCTCTCGCCAAAAAGGCTACCAGGAGCGGGACCTGGAACCCAGTTTTCCATCATCACC
 AAAGTCAAAGGGGTTTCCGTCACTCAGATCAAGGAGCTTGGAAACCGGCTGTGGGATGTG
 GCCGACTTCTGTAAGCCACCTCAGGTGGGGGCTGTGTTGCTGACGGGGGCGCAAGTC
 CTTTCCCCACTGACGCTGAAACACCCGCTATGCAGCCAGTGTGTGCGAGAGAGAAGCAT
 GTGATGCCAGAGACGGCTGCGGGTTCTCAGGAAGGGCTTACAGAGGAGTGGCACCTGGA

3199 ATGTGGCCGACTTCTGTGAAGCCACCTCAGGTGGGGGCTGTGTTGCTGACGGGGCGC
 AAGTCTTTTCCCACTGACAGCCTGAACACCCGCTATGCAGCCAGTGTGTGCGAGAGAGA
 AGCATGTGATGCCAGAGACGGCTGCGGGTTCTCAGGAAGGGCTTACAGAGGAGTGGCAC
 CTGGACAGGACTTTTCAAGGATGTGTAGGAGGTTTGGGGTGGAAAAAGGGGCCACTCAAG
 AAGCCAGGCCAGGGTTGGACGTGCTGGCTCACGCTGTAAATCCCAGCACTTTGGGAGGCC
 [G, A]
 AGGCAGGTGGATCACGAGATTGAGAGTATCTGGCTAACACGGTGAACCCCATCTCTAT
 TAAAAATACAAAAATAGCCGGGCTAGGTGGTGGGCGCCTGTAGTCCCAGCTACTCGGG
 AGGCTGGGGCAGGAGAAATGGCATGAACCCGGGAGGTGGAGCTTGCAGTGAGCCGAGATTG
 CACCACCTGCACCTCAGCCTGGGTGGCAAGCGAGACTCTGTCTCAAAAAAAGAAAAA
 AGCCAGGCCAGAGAACTGCATTTCAAAGACTGCCAACAGAAAAGAGGAGTGTCCAG

3307 GTGCGAGAGAGAAGCATGTGATGCCAGAGACGGCTGCGGGTTCTCAGGAAGGGCTTCACA
 GAGGAGTGGCCTGGACAGGACTTTAGGGATGTGTAGGAGGTTTGGGGTGGAAAAAG
 GGGCCACTCAAGAGCCAGGCCAGGGTTGGACGTGCTGGCTCACGCTGTAAATCCCAGCA
 CTTTGGGAGGCCGAGGCAGGTGGATCACGAGATTGAGAGTATCCTGGCTAACACGGTGAA
 ACCCATCTCTATTAAAAATACAAAAATAGCCGGGCTAGGTGGTGGGCGCTGTAGTC
 [C, G]
 CAGCTACTCGGGAGGCTGGGGCAGGAGAAATGGCATGAACCCGGGAGGTGGAGCTTGCAGT
 GAGCCGAGATTGCACCACTGCCTCCAGCCTGGGTGGCAAAGCGAGACTCTGTCTCAAAA
 AAAAAAAGAGCCAGGCCAGAGAACTGCATTTCAAAGACTGCCAACAGAAAAGAA
 GGGAGTGTCCAGGACTAATGGCTTGGCTTGGAGTGGTGTGAGGTGCTGGGGCATGGAA
 CTTCCCTGTAGCCCTGCTCCCTGACCTGGGGCACTACGGTCAGGTGCTGCTCCTCCCTC

5012 TTAATGACTTGATGGGGCCAACATCCCTTCCCTCATAAACAGGCTGCCGGCTTCCGGCC
 TTTCCAGTCAACAGAGCCAGCCAGGCCAACCTTGAGACTTGCCCTCTAGGGAGAGAAC
 GTGTTCTTCTGGTGACCAACTTCTTGTGACGCCAGCCAAAGTTAGGGCAGATGCCCA
 GAGGTGAGTTTACCCAGGATCCTCCAGCGGGTCCCTTGTTCCTCCATCAGCCCCAGGTG
 GCCACCCGTGTTTCCCTTCCCTTCCAGGTGGCTGAAGGCTCAGCCTGTGCTCGGTGT
 [C, G]
 CCCCAGGCACTGGGCTACATCTTTTCTGAATCATTATGTTTCACTCTTACATATCCCT
 GCCTGGTAGGAAGTCTGTGATCCCATTTTCAAGGAGAGAGTGGGCTCAGTGAGGTT
 GAGTCACTTCTTAAAGGCTTCCAGGCTGTGGGTGACAGGACCCGAGCTCTGGGAGCA
 GCAGTTCATGAGGTGCCAGCCCTCCATCCTGGTCTGCTCTGGGTACTCTCCAG
 GTTGGTAGTGTGACACCCAGAGCTGCGCACATGCTCAGGGAGGTTCTAATAGCAAGAGCC

FIGURE 3, page 7 of 10

6169 CTTGCCTGGGGATGTCCCTGGGATCCTGCATCTGTACAGAGCATGCTCATTCTCTCCAG
CTGTGAATTTTGTGTTGAACTATTGGGACTCAGGACATAGTCTGAAAGTTTACCTCCACA
GTGACATCTTTAGGCAAGTCCAACATTTACGTGCCTCCTGGGCTGGAGGGTCGTTGTGCA
GACAGCTGTCCCTGAGCCCTGGTGGCTGGTCTAGCACAGTTGCTGGAGACATCCCATG
TCCGTAGTTGGAAATATGCACAAAGGATTGCTTACTCTTTTGTGTTTGTGTTTTTTGA
[G, C]
ATGGAGTCTTGCTCTTGTCCTCCCAAGGCTGGAGTTCAATGGCAGCATCTCGGCTCACTGCA
ACCTCCGCCTCCTGGGTTCAAGCAGTTCTCCTGCTCACCCCTGAGTAGCTGGGATTACA
GGTGCCCGCCACTGTGCCAGCTAATTTTGTATTTTAAGTAGAGACGGGGTTTACCAT
GTTGGCCAGGCTGGTCTCGAATCCTGGCCTCAGGTGACCCACCAGCCTCGGCCTCTCAA
AGTGTCTGGGATTACAGGCGTGAGCCTGCCGAGAGCTTGGTCGGGGAGACCTGAACCCAGC

7647 AGGTGGCAGGTCCGATGATGGGACAGAGGCTGTAGGTGGGGACCTAGGGCTGCACTTGA
GCAGAATCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTCTCGCTCTGTACAC
CCAGGCTGGAGTGAGTGGCGTGATCTCGGCTCACTGCACACCTCCACCTCCTTGGTTCA
AGCGATTCTCCTGCCTCAGCCTCCCAAGTAGGTGGGACTACAGGCACACACCACCACT
CGGCTAATTTTGTATTTTAAATAGAGACAGGGTTTTGCTGTGTCGGCCAGGCTGGTCTC
[A, G]
AACTCCTGACCTCAGGTAATCCGCCCACCTTGGCTTCTCAAAGTGTGGGATTACAGGTG
TGCCAGGCCAAGCAGAATCTTAAAAAAGGTGGGAGAGCTGGTGAGCAGGTGGATTG
GTTGAAGCAGGATGTGACACAGAGGGGCTTGGTGGGTAAAGGCCCTGAGCTGTGTGAG
GTGAGGTGCCTTTAGGGCTACCTGCCACTGGGTGGAGCTGAAGTGAAGATTTGAGCTGGG
GTGGAAGAAGGTAGTTTCAAGATTTCAGGGGCCCTGTAAAGCCCCACTAAGGAGCTAAAC

8638 ACAAAGAAGCAGAGCATGTGGCTCTGCTCCGACCTCCACCCAATCACGACGGCCCTGTCT
TTCAGAAAGTCCCACCGCTCATTCTGGCTTCTCAGAGGCCCTCAGCCTTCCTTGCGCC
CTGGTGTCTGTGTTTCTCCTGCTGCCCCCTGAGCTGAGTGCCTGGGAGCAGTGTCCATC
CTCAGTTGGGGCAGGACCATGCCTGGGAGAGTGCCTGATGCTCAAGGGTGCCTTCGTCTC
TGGGCTCTGGGACCCAGAAAGCTCACCTGTCTCCCTTCTGCCAGAGCCCCATAGTCC
[C, T]
ATGCCCTCTGTGAGGCATTAATGTCCCAAGTTACAGAAGAGCGAGCAGGAAGGAGTAGC
CTGTGGTCCCTCAGCAAGGGTGTGGGGTCTGCTTCAATACCAAGCCCTGACTCTAGG
GCCCTGATCTTTGTGAGCTATGTCCCCATGCCGGGCATCAAAAACCTCACCTCCCAAGGT
ATCTTACCTTCCCTGATCTGTATCCAAATTGGACCAGAGGAGCTAGACCTGGAAGAAT
CACTTCCGCATCCACCAGGACAGAACTGTAGGAGGGAAGGGGAGGGTGCCTTGTCTC

9409 TGAGGTGAGAGGATTGCTTAAGCCCGGAGGGCGAGGCTGTAGTGAGCCATGATCATAACC
ACTGCACTAGAGCCTGGACAACAGAGTGAGACCGAATCACTAAAAATAAATTTTGA
AAGGAGAAAGGGGTCTCCCTTTGTCTTTGAAATACAGTACTGTACCTTCATCTGGCCAG
GGCATTGCTCCGCTCCCTCCTCTGACCACCTCCTTTTATTTGACCCCTCCAGCTTCTCTG
TGTGCCCCACACTCAGGGTACTCTGGCGGCGGGTGGTGAGTTGTTTAAAGGTGGGAAG
[T, G]
GGGCTGTCTTCCACCTTGAACCTCCCTGCTTTGAGACTGGGCTGTGGAGGGGAGAC
ATCCCCCTGTGCCATTGGTGACTGCTCTCTCTCCACCTCAGCACCCGTCCGTCCCACTGG
CTAAGTCTGGGTGACGAGGACTGCCCCAAGGGGAGGGAGGCACACAGCCACGGTA
ACTGTGGGCTCTGTCTTCCAGTGCCCTAGCAGGGTGGGGGCCGGGCTGGGATCCTGGGT
GGCTCCTGAGTGACGGCCTGCTCGCCTCTGTCCCTGCATCTCTTTCTGCCAACCAACC

10504 GACCTCGTGGCCAAGGCTGGAGGGACCTTCGAGGACCTGGCGTTGCTGGTGGGTCCCAAG
TTGGGGGAGGGTTCTTAGAGGGCTCTGGGAGAGGGTCCCGGGCCACCCACCGGTGGAA
AAGCTATGTGCTATGTGAGGGTGGCTCTGTAGGCATCAGAGTTTACTGGGATTGTGACC
TGGACACCGGGGACTCTGGTGTCTGGCCTCACTACTCCTTCCAGCTGCAGGAGAAGAGCT
ACAACTTCAAGTGAGGCCCACTGCTCCAGTGCCAGCTGCTGGGCCCATCGCCCTCTC
[A, C]
CTGTGGCGGCCAGGACAGACACACCCAGGCCAGGCTCTAGATATTCACCTACGTGTG
CAAGGGGGTCCCAGGAGCAGGAGAGAGCTGTTCTCAACCCACATCCTCCAGCACAGGCT
CCGTCTGTGCCCCAAGTCTGAGCCCTCCACCCCATCTGTCCAGGCCCTTCCCGAGC
TCAGGCTCTCTCACTGCCAGCCCTTCTCCACCCACCTCGCTTCTAGTATCTCCCTCCA
CAGCAATGGGGTGTTCATTTTTTACTTTCCCTTCTCCCTTCAGCTTGTGTTTTTTTTT

10971 GGCCCTGCCAGCTCAGGCTCCTCACTGCCAGCCCTTCTCCACCCACCTCGCTTCTA
GTATCTCCCTCCACAGCAATGGGGTGTTCATTTTTTACTTTCCCTTCTCCCTTTCAGC
TTTTTTTTTTTTTTTAAAGACAGAATCTCATCTGTACCCAGGCTGGAGTGAGTGGC
CCGACCTCGGCTCACTGTAACCTCTGCTTCCCTGGGTCAACCGATTCTCCTTCCCTCAGCC
TCCTGAGTAGCTGGAATTACAGGTGCTCGCCACTACTCCAGCTAATTTTTATATTTTGG
[T, -]
AGATAGAGATGGGTTTTTCAATGTTGGCCAGGCTGGTCTCAAACCCCTGACCTCAGGTG
ATCCACCCACCTCAGCCTCCCGAAGGGCTAGGATTACAGACGTAAACCCATGCTGCGC
CTCCCTTCCGCTTTTACCTAAACTTTTTTTTTTTTTTGGAGATGGAGTCTCACTCTGTG
CCAGGCTGGAGTACAGTGGCGGATCTCAGCTCACTGCAAGTTCCGCTTCCCGTGTTC

FIGURE 3, page 8 of 10

CGCCATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTACGGGTGCACGCCTCCACGCC

12609 CCAGGAAGTCTACCCAGTTCCAGGGAAGAGTGAGTTCATCTCTGGAATCCCTCAGC
CCTGAGCCTGCCCCCTTCACATCCCCGCTGCTGGGTCTGTTTAGGGACTCCTCTGTCCCC
CGTCTCTCAGCAGGCAGGGAAGTCTGAGGGACAGGTCTTCGTTTGTCTTTTCTGTTTT
CTCACCAATTACATAGGCTGAGACCCAGGACTCAGGCTTGGGCTGGGGGTTTATAGAGT
CAATTGACAAGTTGGACAGAGTCTGGCAGGGCCAGCCCCACCTGGGGTGGGCAAAGCA
[G, A]
GTCACCAGAGCCTTCTTTCTGCCCACAGGACAGCCACTCACTGGTGGGAGCAACCGGGT
GTGAGGCCCCGACCCCTGCTCAAGCTCTATGGAATCCGCTTCGACATCCTCGTCACCGGG
CAGGTAGGCACAGGTAGGGGTGAGGCCGGGATGGGATGGGGCAGGCAGACAGGGCTGGA
GGAGGCATGAGGCTGACAGTCTGTTGGCTGAGAGGTTGAGCTCAGATCTCTCTCAGGCAGG
GAAGTTCGGGCTCATCCCCACGGCCGTCACACTGGGCACCGGGGAGCTTGGCTGGGCGT

13367 TTGGCCCTGCCTCTCCAGGTACCTTTTTCTGTGACCTGCTACTGCTGTATGTGGATAG
AGAAGCCCATTTCTACTGGAGACAAAGTATGAGGAGGTGAGCTGAGGTGCTCTGCTTG
GACCTGGGTTCTGCCACACTTAGGAAGATGTTGGCTGGATCCCTGACCTGCTGTCTCA
TCTGCAGGCCAAGGCCCGAAAGCAACCGCAACTCTGTGTGGAGGGAGCTGGCCCTTGC
ATCCCAAGCCGACTGGCCGAGTGCCTCAGACGGAGCTCAGCACCTGCACCCACGGCCAC
[T, A]
GCTGCTGGGAGTCAGACACAGACACCAGGATGGCCCTGTCCAAGTTCTGACACCCACTTG
CCAACCCATTTCCGGGAGCTGTAGCCGTTCCCTGCTGGTTGAGAGTTGGGGGCTGGGAAG
GGCGGGGCCCTGCCTGGGGATCTCAAGGATGAGGCCCCAGCATGGAGGATTGGGGGTAGA
ATTCACCCCTGAACCCAGCAGACAGTCCCTCCCTGACTCCCACTTGGTAGGGTGCT
GCCTCAGGGAGCCATAGAAGTCGGCTGTGTTTTGAGACGGCGACAGAACCTGACCCGTGG

14191 ATCGGTCTCATATGGGGCTGTGCAGCTGGAGCCAAAAGGCAAGGTAGAAAGAGGAGTGA
TGGGGGAGGGGATGTTTCAGCTTCTCTGGTGCTGTGATGCCCCAGGAGAGTCTTAATC
TAGGGAATGGGGTGGAGTAGGCAGATAATCCACCTCCCTATCCCCCAGGCAAGGGCGGAG
CATGTGCTTGGGCCCAACCTGCTTAGTTTATGAGGACCGGCTGCTTTCCAGTGGTAGC
CCTTTTGCCATGGAGGTCTGGGAGAGAGAGCAGAGGGCGGCAGGGCTAAGTTGGTGATCA
[T, C]
TGGGTTCTTCAGGACCTTCTATATCCCTCCTCGGTAACCCCCAGGCCCAACCCCTTGGA
TCTTTCTCCAGGCTTCTGAGAGCCCTGGGGGTGGGAGGCTGTGGGAGGCTGTACATCT
GAAATTCACCTTCAGTCCAAGTCATACCTAGGAAGCTGTCTGGGCAGCTGCTCGAGGGAGG
CCCTGGGCTCTGATCCAGGCTGGATGGAGTGGCTGGAAGGAATGGTCCAAACAACACCA
CCGAGATCTCCCTCAGGCTGGCCAGGTTTTCAGCTGGAATTCTCCTCTTGGTCCCAGGG

14227 AAGGCAAGGTAGAAAGAGGAGTGTATGGGGGAGGGGATGTTTCAGCTTCTCTGGTGCTG
TGTATGCCCCAGGAGAGTCTTAATCTAGGGAATGGGGTGGAGTAGGCAGATAATCCACCTC
CCTATCCCCCAGGCAAGGGCGGAGCATGTGTCTTGGGCCACACCTGCTTAGTTTATGAG
GACCGGCTGCTTTCCAGTGGTAGCCCTTTTGCCATGGAGGTCTGGGAGAGAGAGCAGAGG
GCGGCAGGGCTAAGTTGGTGATCATTGGGTTCTTCAGGACCTTCTATATCCCTCCTCGGT
[A, G]
ACCCCCCAGGCCAACCCCTTGAATCTTTCTCCAGGCTTCTGAGAGCCCTGGGGGTGG
GAGGCTGTGGGAGGCTGTACATCTGAAATTCACCTTCAGTCCAAGTCATACCTAGGAAGCT
GTCTGGGCAGCTGCTCGAGGGAGGCCCTGGCTCTGATCCAGGCTGGATGGAGTGGCTGG
AAGGAATGGTTCCAAACAACACCCAGGATCTCCCTCAGGCTGGCCAGGTTTTGACGCT
GGAATTCCTCTTGGTCCCAGGGCGGGCAGGGAATCTAAGTGTCCACCCAGGGAGG

15027 AGGGCCCCTGAGGCTGGGTATCCAAGGAGGGGCAGTGCACCTGATTCTCCTTGGGGCC
CAGAGGAAGCTGATGTATGGCTGGACAAAGTCACGGAGTAAAGCCAGCAAAGCCACCT
CTTCTGTGTAGTCTTACAGGCATGACTGGAAAGTTGGGGGCATCTATGGTAGACATG
GCACAGCCATGAAGAGACCAGTGGGGTGGTGCAGGGTGGACTTGGGGACCTACCCCTGA
AGACTGAGGCCCTGCAGTACCAGGTGGGCTAGAAGGTAAGTGAACAGGCCTGGGCACT
[T, C]
GTGCACCCATGTAGGAGCATGAGGGCCACACTCTTTTCACTCAAAGCCCTGAAGAGTG
GGCAAAGACAGCAAGAGAGCTGCAGCCTGGGCCCAGCTCAGAAACAGCTGTGCGCTCAG
TCTGCGCACAGGCATGCACCCAGGGTAGTGCTGCAGGGATGCATGTGTCCCGTGGGG
GTGCTGTGCCAGGCAGGCCCTCAGGTGCATGCCATGCTCAGAACCCTGTGCCCCCTTCTA
GGCAGCCTCCTTGGGGCCAAGCTCTGCTCCCTGGATCTGCCACCTAGCAGACGTGGGGA

FIGURE 3, page 9 of 10

15441 GCCTCAGTCTGCGCACAGGCATGCACCCAGGGTAGTGCCTGCAGGGATGCATGTGTCCC
 CGTGGGGGTGCCTGTGCCAGGCAGGCCTCAGGTGCATGCCATGCTCAGAACCCTGCTGCC
 CTTTCTAGGCAGCCTCCTTGGGGCCCAAGCTCTGCTCCCTGGATCTGCCACCTAGCAGAC
 GTGGGGAGCCTGACCCCATGCCTGTGCATGGAACCCTCCTTGCCTGGTGTGTGTGGCTCCC
 CTCTTCACTGGGCACCTGGATCCAGGCCACCTGTGTCCCTGACTCAGGGTGGTCCCAGG
 [A, C]
 CTGGCACCTACTCTTTAGAGAGCCCCAGCATCTTTGATGTGGATTGGAGACAATTGCCTG
 GTTCCCTGGGGCAGGTGAAGACTTGGTGCCACAAGAATGCCACAGTGGATACGCCAGCA
 GGCCACATGGCTGGCCAAGCAATTATTATTATGGATCCCTTGGGCTGTGGGCCTTCCCAT
 CCACCCCAACCAACTGCCAGGTAGCTGGAGCTGATCATAACAAGAAGGCTCTGGGCA
 GAGTCCATGGCACCAGCACCAGCCAAGGCCCACTCCTGAAGACCCGAAGCCAGCCCCTG

Chromosome map:
 Chromosome No: 22

15441 GCCTCAGTCTGCGCACAGGCATGCACCCAGGGTAGTGCCTGCAGGGATGCATGTGTCCC